

VU Research Portal

Plants, microbes and markets

Werner, G.D.A.

2016

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Werner, G. D. A. (2016). *Plants, microbes and markets: Evolution and maintenance of belowground cooperation*. [Vrije Universiteit Amsterdam]. Uitgeverij BoxPress.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

CHAPTER 8

Evolutionary signals of symbiotic persistence in the legume-rhizobia mutualism

Gijsbert D.A. Werner, William K. Cornwell,
Johannes H.C. Cornelissen and E. Toby Kiers

Published in Proceedings of the National Academy of Sciences of the
United States of America (2015) 112(33): 10262-10269

Abstract

Understanding the origins and evolutionary trajectories of symbiotic partnerships remains a major challenge. Why are some symbioses lost over evolutionary time, while others become crucial for survival? Here, we employ a quantitative trait reconstruction method to characterize different evolutionary stages in the ancient symbiosis between legumes (*Fabaceae*) and nitrogen-fixing bacteria, asking how labile is symbiosis across different host clades. We find that more than half of the 1,195 extant nodulating legumes analysed have a high likelihood (>95%) of being in a state of high symbiotic persistence, meaning they show a continued capacity to form the symbiosis over evolutionary time, even though the partnership has remained facultative and is not obligate. To explore patterns associated with the likelihood of loss and retention of the N₂-fixing symbiosis, we tested for correlations between symbiotic persistence and legume distribution, climate, soil and trait data. We found a strong latitudinal effect and demonstrated that low mean annual temperatures are associated with high symbiotic persistence in legumes. While no significant correlations between soil variables and symbiotic persistence were found, nitrogen and phosphorus leaf contents were positively correlated with legumes in a state of high symbiotic persistence. This pattern suggests that highly demanding nutrient lifestyles are associated with more stable partnerships, potentially because they ‘lock’ the hosts into symbiotic dependency. Quantitative reconstruction methods are emerging as a powerful comparative tool to study broad patterns of symbiont loss and retention across diverse partnerships.

Introduction

Symbiotic partnerships have transformed the Earth's nutrient cycles (Vitousek *et al.*, 2013), driven the diversification of organisms (Joy, 2013; Weber & Agrawal, 2014), and facilitated rapid adaptation of species to divergent new niches (Moran, 2007). Almost all organisms rely on symbiotic associations for some form of metabolism, protection or energy (Bronstein, 1994; McFall-Ngai *et al.*, 2013). These relationships can be stabilized if the interests of the partners are tightly aligned or partners have evolved mechanisms to control and coordinate interactions (Aanen *et al.*, 2009; Kiers *et al.*, 2011; McFall-Ngai, 2014). However, symbioses can also be unstable over evolutionary time (Toby *et al.*, 2010; Sachs *et al.*, 2011b; Ghoul *et al.*, 2014), and conflicts among partners can lead to breakdown of potentially beneficial relationships, and even drive the emergence of parasitism (Sachs & Simms, 2006; Merckx & Bidartondo, 2008). Although our theoretical understanding of the factors that promote cooperative relationships has increased (Foster & Wenseleers, 2006; Palmer *et al.*, 2010; *e.g.* Leigh, 2010; Golubski & Klausmeier, 2010), we have lacked comparative approaches that can be used across a diversity of symbiotic systems to ask why some partnerships persist over millions of years and others are more labile. What distinguishes host species that acquire highly stable symbioses from those species that do not? Do different hosts follow similar trajectories of symbiotic evolution? Can different stages of symbiotic evolution be correlated to specific ecological conditions? One tool to study broad patterns of symbiont evolution is quantitative trait reconstruction (Felsenstein, 2012; Beaulieu *et al.*, 2013; Revell, 2014), a comparative method that allows us to (1) identify when symbioses become more or less persistent and (2) correlate these levels of persistence to ecological factors and partner traits.

Persistent symbioses

We define symbiotic persistence as the continued capacity of an organism to form a particular symbiosis over evolutionary time (*e.g.* millions to tens of millions of years). High symbiotic persistence means that the capacity to form the symbiosis is more likely to be retained in a species or lineage, whereas highly labile symbioses are characterized by low persistence and frequent loss of symbiotic capacity. As defined here, symbiotic persistence is an inherently phylogenetic concept (*i.e.* increased evolutionary stability of symbioses), meaning that species with an increased symbiotic persistence may be phenotypically indistinguishable from more labile symbioses. Symbiotic persistence is not the same as a symbiosis being obligate or genetically integrated. Although we expect intracellularly integrated partnerships to typically have high symbiotic persistence, the reverse is not necessarily true. For instance, some symbioses successfully persist for millions of years, without ever integrating intracellularly into the host (Remy *et al.*, 1994; Currie, 2006; McFall-Ngai, 2014; Kaltenpoth *et al.*, 2014; Kremer *et al.*, 2014).

When a symbiosis has become fully persistent the loss rate falls to zero, or nearly zero as even some usually presumed permanent symbioses (*e.g.* mitochondria or chloroplasts in multicellular organisms) can be lost in extremely rare cases (Merckx & Freudenstein, 2010; Hjort *et al.*, 2010).

The level of symbiotic persistence can be quantified using modern comparative methods, allowing us to estimate the likelihood that a symbiosis will be retained, rather than lost (Beaulieu *et al.*, 2013; O’Meara & Beaulieu, 2014). Loss rates across phylogenetic trees can be compared to identify species or clades associated with more persistent symbiotic partnerships. We can then attempt to correlate highly persistent symbioses with global distribution maps, climate and soil databases or organismal trait databases to elucidate the environmental conditions and species characteristics associated with evolutionary persistence.

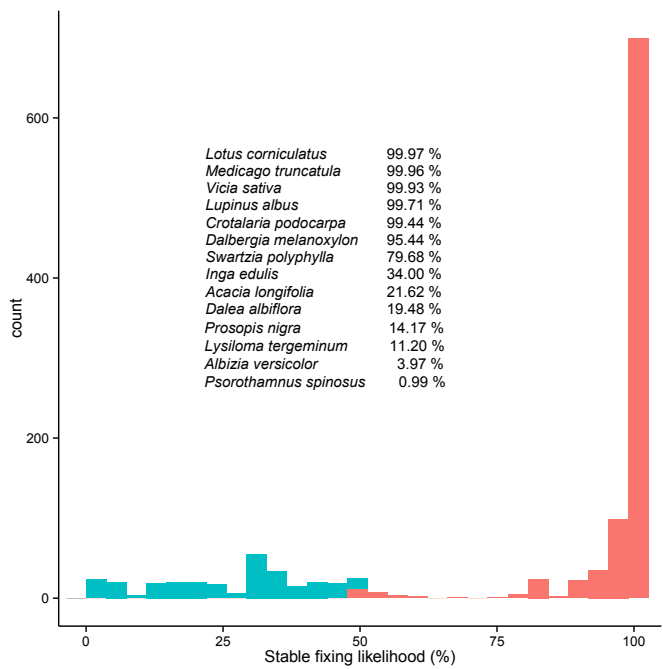


Figure 1: Histogram of the stable fixing likelihood for all nodulating legume species in our reconstruction. To illustrate this variation, we list a range of species and their associated likelihoods of being stable fixers. Species with a stable fixing likelihood of over 50% are labeled red, the other species are labeled blue.

Testing for persistence: the case of the N₂-fixing symbiosis

Our aim is to quantify symbiotic persistence in the mutualism between nodulating plants and their nitrogen-fixing bacteria. We focus specifically on plants in the *Fabaceae* (legumes) associating with a class of nitrogen-fixing bacteria known broadly as rhizobia (Young & Haukka, 1996). In exchange for atmospheric nitrogen fixed by bacterial partners, plants provide rhizobia with photosynthate for energy and highly specialized root structures, called nodules, to house bacteria (Sprent, 2009). Although rhizobia can survive and reproduce in the soil, colonisation of a legume and establishment of a successful symbiosis can lead to a 10⁸-fold increase in a single rhizobial cell's abundance (Denison & Kiers, 2011), thereby creating a strong selection pressure for symbiotic engagement. Host plants can also survive in the absence of bacterial partners but do, in many ecological situations (Vitousek *et al.*, 2002), benefit greatly from the nitrogen the bacteria provide (Thrall *et al.*, 2000; Lee *et al.*, 2003).

The legume-rhizobial symbiosis is a good case study to quantify symbiotic persistence for a number of reasons. First, the phylogenetic relationship among potential hosts is well characterized (Zanne *et al.*, 2014) and a newly compiled database of which hosts engage in the rhizobial symbiosis is available (Werner *et al.*, 2014b, a; Chapter 7). Second, the symbiosis is found across a range of plant hosts, spanning highly diverse habitats and ecosystems (Doyle, 2011; Sprent *et al.*, 2013), allowing for potential variation in symbiotic persistence and in the selection pressures driving it. Third, different morphological types of the symbiosis have evolved across host plants (Sprent, 2009; Masson-Boivin *et al.*, 2009; Sprent *et al.*, 2013), suggesting that there is potential for variation in the evolutionary trajectory of the symbiosis. Lastly, because the symbiosis is facultative for both partners, it represents an interesting case in which high symbiotic persistence is not necessarily linked with strict mutual dependence or host-symbiont integration, as is the case with eukaryotic organelles or some insect endosymbionts (Van Leuven *et al.*, 2014; Bennett & Moran, 2015).

We previously used a phylogenetic reconstruction approach called “Hidden Rate Models” (HRMs) (Beaulieu *et al.*, 2013; O'Meara & Beaulieu, 2014) to study the evolutionary history of the N₂-fixing symbiosis across the angiosperms (Werner *et al.*, 2014a). These models allow for heterogeneity in the loss rate of a trait, like a symbiotic partnership, and thus can identify the phylogenetic signal where host clades shift in and out of different evolutionary states (Beaulieu *et al.*, 2013; O'Meara & Beaulieu, 2014), such as when labile symbiotic associations become more stable. In our previous work we applied such a model to 3,467 angiosperm species, 1,709 of which were in the four symbiotic nitrogen-fixation orders (*Fabales*, *Fagales*, *Rosales* and *Cucurbitales*), known as the nitrogen-fixing clade (Soltis *et al.*, 1995; Doyle, 2011). We found that across the nitrogen-fixing clade, symbiotic host plants can be in two different evolutionary states:

a ‘regular fixing state’ where the ability to form the symbiosis remains highly labile (1.17 losses per 100 million years per lineage) or a second class termed ‘stable fixing state’ in which host plants are characterized by extremely low loss rates (0.02 losses per 100 million years per lineage), *i.e.* almost 60 times lower than that of hosts in the ‘regular fixing state’ (Werner *et al.*, 2014a). The analysis suggested that for plants in a state of stable fixing, it is almost impossible to lose the capacity to form the symbiosis (Werner *et al.*, 2014a).

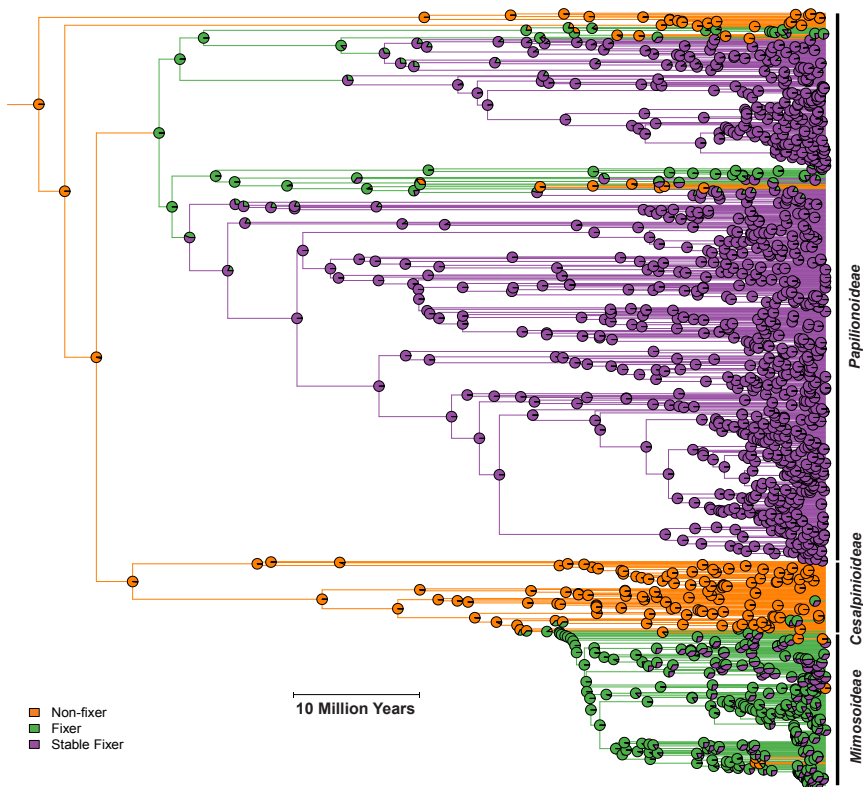


Figure 2: Pie charts indicate the likelihood of a host node being in either a non-symbiotic state (orange), a ‘regular’ N_2 -fixing state (green), or in a stable fixing state (purple), where the symbiosis has a high persistence.

Represented are the legumes from their ancestral node about 62 million years ago. Following the evolution of symbiotic nitrogen fixation (transition from orange to green), repeated evolution of stable fixing (transition from green to purple) is found across the legumes. Particularly within the *Papilionoideae*, legume species are very likely to currently be stable fixers, meaning they show a continued capacity to form the symbiosis over evolutionary time. Scale bar indicates 10 million years.

What plant hosts have entered into this stable symbiotic state and why? Can specific ecological factors and host traits be correlated with the stable fixing state? To answer these questions we further studied the 1,366 legume species in our analysed dataset, of which 1,195 are capable of forming the symbiosis (Werner *et al.*, 2014b). Legumes are the best characterised order in the nitrogen-fixing clade and have a wide geographic range (Doyle, 2011; Sprent *et al.*, 2013), ensuring we could explore a range of potential environmental conditions and host traits, and there is considerable variation in their ability to successfully form a symbiosis with rhizobia (Sprent, 2009). We first used an HRM approach to test if the best fit model of symbiosis evolution in angiosperms was similar to that in legume species. To establish if this model still has the best fit, we repeated our HRM-analysis using the legume dataset and the same phylogeny, this time pruned to the legumes (Zanne *et al.*, 2013, 2014). We found that that in agreement with the angiosperm analysis (Werner *et al.*, 2014a), our best model included both a regular fixing state and a stable fixing state with increased symbiotic persistence (89.6% AIC-weight). Additionally, we found a highly significant correlation between the stable fixing state likelihoods of extant nodulating legumes species from our previous analysis (Werner *et al.*, 2014a) and our analysis of legumes only ($r = 0.92$, $p < 0.01$). These observations confirm that our model of symbiosis evolution across the angiosperms also provides an accurate representation of the evolution of stable fixation within the legumes.

We then mapped our calculated stable fixing likelihoods (Werner *et al.*, 2014b) (Figure 1) onto the legume phylogeny (Figure 2) to analyse the evolutionary trajectory of the stable fixing state and to explore its correlation with ecological factors and host traits. The first observation is that there is considerable variation in symbiotic persistence. High stable fixing likelihoods (>95%) were found in 803 of the symbiotic legume species studied, whereas 95 symbiotic species had very low likelihoods (<20%) of stable fixing (Figure 1). The second observation is that the evolution to the stable fixing state is likely to have multiple origins across the legumes (Figure 2). This finding suggests that there was no single event or innovation that allowed hosts to transition into a stable state, but rather there were several parallel routes towards persistent capacity for symbiosis (Werner *et al.*, 2014a). Third, we observed that within the *Papilionoidea*, there are repeated likely origins of stable fixing (Figure 2). These origins range from the oldest over 50 million years ago, to a more recent origin a little over a million years ago, in the genus *Amorpha* (false indigo). This observation illustrates how, even within single clades, there is large variation in the timing of symbiotic stabilization. Fourth, we see that not all legumes have reached or are close to reaching the stable state (Figure 1) which emphasizes that in many legume clades, like in the *Mimosoideae* (Figure 2), the symbiosis is easier to lose.

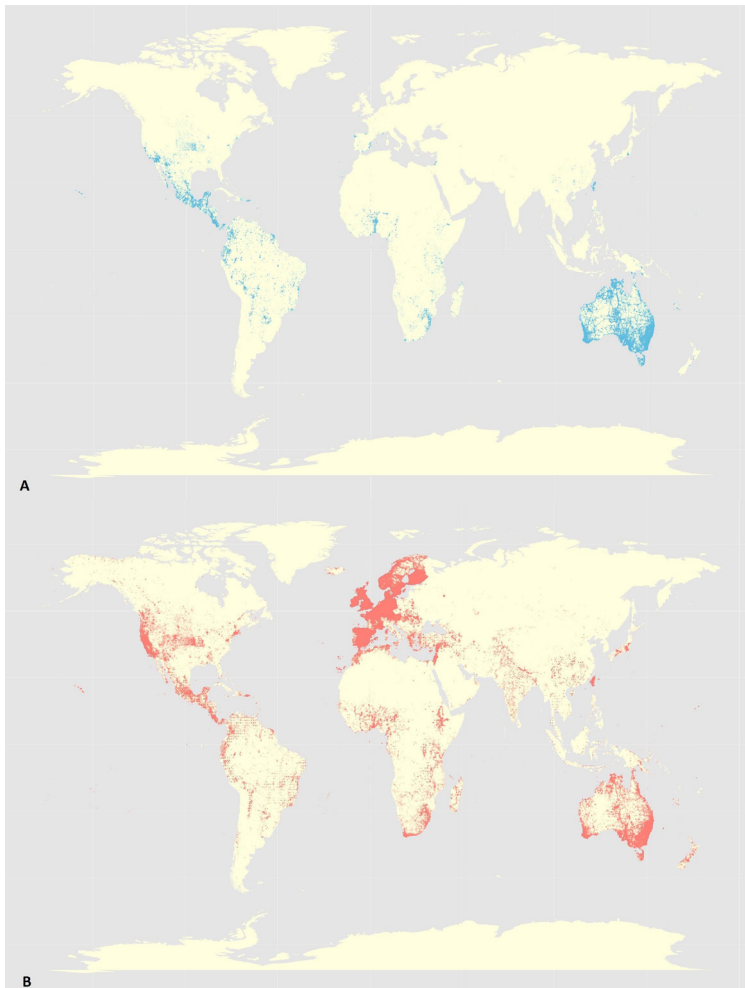


Figure 3: Occurrences of regular (blue) and stable fixing (red) legumes across the globe extracted from Global Biodiversity Information Facility. Over 3.2 million occurrence records were plotted for the nodulating legumes analysed in our dataset (Methods).

Correlations between symbiosis persistence and ecological factors or host traits

Quantitative reconstruction methods, such as the hidden rate models (Beaulieu *et al.*, 2013; O'Meara & Beaulieu, 2014), allow researchers to quantify the likelihood across the host phylogeny of evolutionary states that cannot necessarily be identified through phenotypic differences, such as the stable fixing state, and then ask if there are ecological or trait commonalities among hosts in the stable fixing state. Specifically our aim was to ask how occurrence, climate, environment and host trait data can be related

to particular evolutionary states. We characterized symbiotic persistence in nodulating legumes by testing potential explanatory factors at four levels: (i) global geographical distribution, (ii) climate variables, (iii) local soil variables and (iv) plant traits. For all variables considered, we explored the correlation of the potential explanatory variable (distribution, climate, etc) with the underlying continuous likelihood of being a stable fixer as the response variable (Werner *et al.*, 2014b,a).

Latitudinal pattern in symbiotic persistence is driven by temperature

We first asked if there were differences in global distribution between stable fixing legume species (>50% stable fixing likelihood) and legumes where the symbiosis is more easily lost (<50% stable fixing likelihood, Figure 1). We used the Global Biodiversity Information Facility (GBIF) to obtain (after quality control steps) a total of 3.2 million legume occurrences for 1,182 nodulating legume species in our analysis (Chamberlain *et al.*, 2014; Methods). We plotted the occurrence of both regular (Figure 3A) and stable fixers (Figure 3B) onto the map of the world. We found that in tropical regions, both regular and stable fixers are common. In contrast, in temperate, and particularly in polar regions, the legumes are almost exclusively characterized by highly persistent symbioses. When we considered the absolute median latitude of fixing legume species as potential explanatory variable for stable fixing likelihood, we found a significant positive correlation using phylogenetic regression to correct for phylogenetic non-independence (coefficient: 0.04% increase in stable fixing likelihood per degree absolute latitude, $p = 0.02$) (Pagel, 1999; Tung Ho & Ané, 2014).

We then asked if the correlation between latitude and symbiotic persistence could be explained by mean annual temperature (TMP_{Mean}). For each legume occurrence, we extracted the annual mean temperature (Figure 3) from a global interpolated climate grid (New *et al.*, 2002), obtaining a total of over 1.4 million temperature records, and calculated the median species-level TMP_{Mean} over all the locations of a species for the 1,095 species for which sufficient data were available (Methods). We then used AIC-criteria to test models that included temperature, latitude, and their interaction. We found that the only significant term was a negative correlation between the median annual mean temperature and stable fixing ($p = 0.04$), and that the previously significant latitude term was no longer significant in a full model containing both terms (phylogenetic regression ($p = 0.41$)). The negative correlation between temperature and stable fixing was confirmed when we simplified the model to only include the significant model term temperature (coefficient: -0.1% change in stable fixing likelihood per degree TMP_{Mean} , $p = 0.01$). This finding suggests that the observed latitudinal pattern of stable symbiosis distribution is driven by temperature.

Soil properties are not correlated with symbiotic persistence in legumes

We then considered local environmental variables, specifically focusing on soil characteristics. We were interested in whether stable fixing hosts were positively correlated with low resource soils, particularly N availability. We extracted soil nitrogen, soil pH and clay content based on our legume occurrence data (Figure 2) from the Harmonized World Soil Database (Nachtergaele *et al.*, 2012) and the ISRIC-WISE database (Batjes, 1995, 2012), obtained a total of over 1.2 million records per soil variable and calculated species-level median values for 1,095 species (Methods). Using AIC-criteria, we tested various models, which included the soil factors and their interactions. We found that none of the three soil variables improved the fit of our phylogenetic regression model compared to a model that only fits the intercept. We therefore rejected soil N, clay content, and pH as factors correlated with symbiotic persistence in the N_2 -fixing symbiosis in legumes.

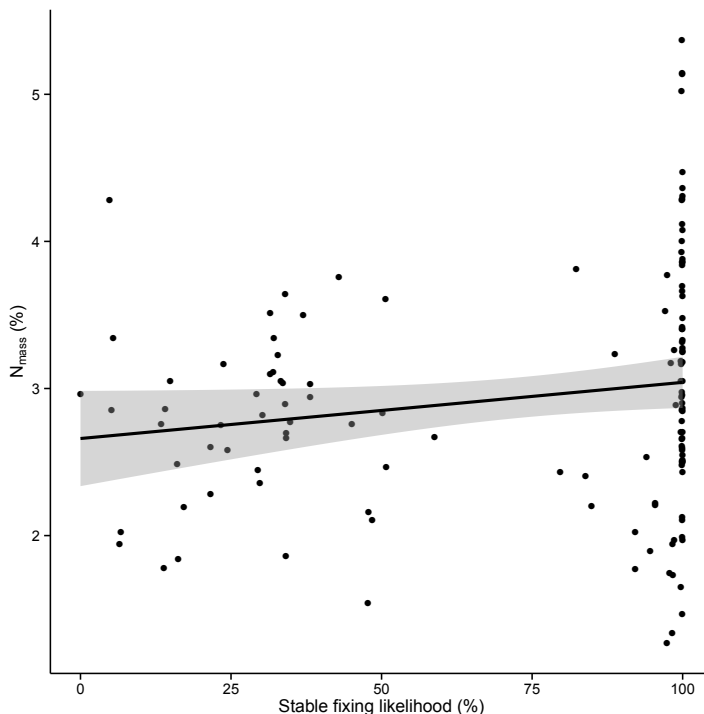


Figure 4: Positive correlation between the calculated likelihood of extant nodulating legumes being in a stable fixing state and the average leaf nitrogen content of these same species (phylogenetic regression, $p = 0.03$)

High host phosphorus and nitrogen contents correlates with symbiotic persistence

Lastly, we tested for correlations between stable fixing state likelihoods and host plant traits, focusing on leaf nitrogen and phosphorus content, asking whether the stable fixing state was associated with greater accumulation of nitrogen and phosphorus (McKey, 1994; Vitousek *et al.*, 2002). We obtained plant leaf nitrogen content (N_{mass} %; 1,114 records) and phosphorus content (P_{mass} %; 612 records) and calculated averages for 131 fixing legumes from our phylogenetic analysis for which data were available (Methods). We tested the association between stable fixing and host plant nutrient content in a phylogenetic regression model, and found positive and significant relationships between N_{mass} % (coefficient 0.03, $p=0.03$; Figure 4) and P_{mass} % (coefficient 0.05, $p=0.03$; Figure 5) and the likelihood that a species is currently a stable fixer, as well as a negative interaction between N_{mass} % and P_{mass} % (coefficient: -0.01, $p = 0.04$). These positive correlations suggest that a higher symbiosis persistence is associated with a higher average leaf nitrogen and phosphorus content in legume hosts whereas the negative interaction term suggests that the positive relationship of either nutrient with symbiosis persistence diminishes with increases in the other nutrient.

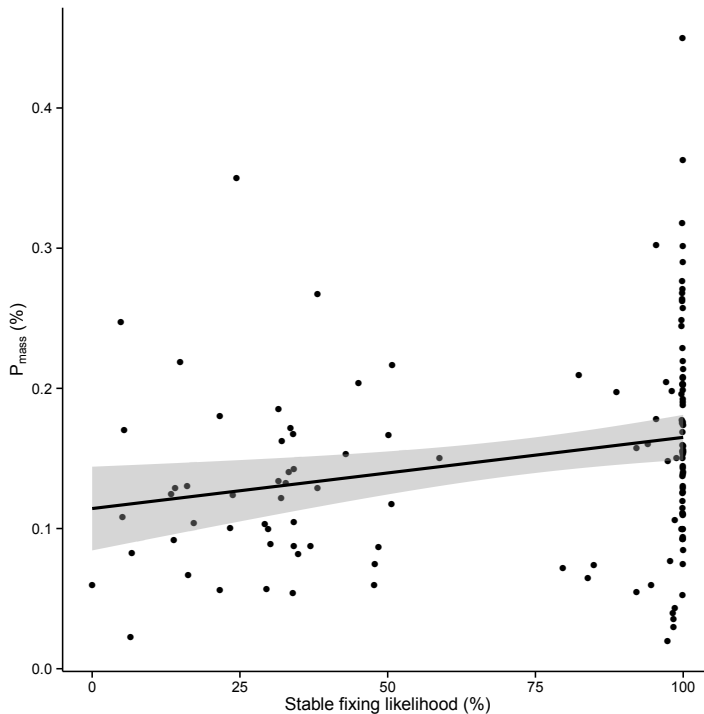


Figure 5: Positive correlation between the calculated likelihood of extant nodulating legumes being in a stable fixing state and the average leaf phosphorus content of these same species (phylogenetic regression, $p = 0.03$).

Discussion

Symbiotic associations can be highly labile, with partnerships gained and lost over relatively short evolutionary time periods. Other symbioses show remarkable persistence over hundreds of millions of years. Although a suite of genomic tools has emerged to study symbiotic persistence as the physical process of host-symbiont integration (de Duve, 2007; McCutcheon & Moran, 2012; Husnik *et al.*, 2013), we lacked an approach that allowed researchers to ask broader-scale questions about patterns and processes across diverse symbiotic systems. Quantitative phylogenetic reconstruction methods can be used to first identify clades with an increased symbiotic persistence, and then to analyze factors that can be correlated with persistence likelihoods. Studying symbiotic persistence in the legume-rhizobia mutualism is interesting because: (i) there are legume clades in which the symbiosis is relatively labile over evolutionary times and others where it is quite stable (Figures 1 and 2) and (ii) the symbiosis is not linked to mutual dependence or to intracellular integration (McCutcheon & Moran, 2012; Husnik *et al.*, 2013; Bennett & Moran, 2013), despite host benefits of N₂-fixation in many ecological conditions (Thrall *et al.*, 2000; Lee *et al.*, 2003). These findings stress that not all highly persistent symbioses are on a route to intracellular lifestyles (Kiers & West, 2015; West *et al.*, 2015). What factors determine if a symbiosis is maintained or easily lost on an evolutionary time scale?

We performed a phylogenetic analysis of symbiotic persistence in legume hosts to explore possible geographic, climate, soil variables, and host traits that might correlate with symbiotic state likelihoods in legumes. While research has demonstrated global variability in symbiosis investment, for example physical intensity of host root colonisation by symbiotic mycorrhizal fungi (Soudzilovskaia *et al.*, 2015), there has been no attempt to map global patterns of evolutionary states of a symbiosis. First, we analysed the geographical distribution of stable fixers across the globe and found a clear pattern showing that at higher latitudes, legumes are generally stable fixers (Figure 3A). In contrast, closer to the equator, the symbiosis can be highly persistent in some clades, as well as labile in others (Figure 3B). These observations have an interesting parallel with a recent model that showed a latitudinal shift in N₂-fixing tree strategies, with a constant rate of nitrogen-fixation being favoured at high latitudes and a plastic, flexible rate closer to the tropics (Menge *et al.*, 2014). We found evidence that the latitudinal pattern of stable fixing is driven by temperature differences among species occurrences, with low annual temperatures associated with high symbiotic persistence. While we cannot exclude the possibility that this pattern is driven by an unmeasured variable associated with global temperatures, this result is further confirmed by the relatively high occurrences of stable fixers in other low-temperature habitats at lower latitudes, such as the Himalayas and the Andes (Figure 3B). An interesting hypothesis to consider is that the negative relationship between symbiosis persistence and temperature is linked

with higher rates of molecular evolution in the tropics versus higher latitudes (Wright *et al.*, 2006; Gillman *et al.*, 2009). Experimental work has revealed that mutations in single genes can be sufficient to disrupt a legume's capacity for symbiosis (Heckmann *et al.*, 2006; Soyano *et al.*, 2013). Although speculative, higher rates of molecular evolution driven by temperature could disrupt key gene networks involved in nodulation, thereby driving loss of the capacity for symbiosis. A similar positive effect of temperature on symbiosis loss has been reported experimentally in rhizobial bacteria (Zurkowski, 1982). Whether this relationship between low temperature and high persistence holds for other symbiotic systems requires further exploration, but illustrates how patterns emerging from quantitative reconstructions can be used to answer broad questions about symbiosis evolution and stability. Another potential explanation for the observed latitudinal pattern relates to relative phosphorus and nitrogen limitations to plant hosts. Colder biomes are thought to be relatively nitrogen-limited, especially in formerly glaciated regions where time for nitrogen to accumulate from biotic N_2 -fixation has been relatively short (Shaver & Chapin, 1995; Shaver *et al.*, 2001; Aerts *et al.*, 2006; Lambers *et al.*, 2008; Pérez *et al.*, 2014). In contrast, the (sub-)tropics tend to have very old, weathered soils depleted in phosphorus (Lambers *et al.*, 2008, 2009; Turner *et al.*, 2013), where special adaptations to acquire phosphorus are presumably more important than N_2 -fixation. Such differences in soil type could favor the evolution of stable N_2 -fixing symbioses in polar regions.

Linking persistence of the N_2 -fixing symbiosis to N/P ratios remains an open question. However, we were able to extract three soil properties for our global legume occurrences to test the idea that stable fixing is associated with lower quality soils. We expected to find increased symbiotic persistence when environmental nitrogen availability, a function of soil pH, clay content and nitrogen content (Shaffer *et al.*, 2001), was lower because hosts theoretically derive more benefit from symbiosis in low-quality environments (Hochberg *et al.*, 2000; Thrall *et al.*, 2007; Johnson, 2010). Surprisingly, we found no significant relationship between soil variables and symbiotic persistence. This result could have a few potential explanations. The first is that most legumes are capable of down-regulating, and temporally controlling nodulation in high soil nitrogen conditions (Müller *et al.*, 1993; Batterman *et al.*, 2013b). Such control means that legumes can limit the potential negative effects of engaging in symbiosis in high nitrogen conditions, and thus might not experience strong selection to lose the capacity for nodulation. However, this scenario assumes there are few costs associated with hosts maintaining the symbiosis, which is unlikely (Menge *et al.*, 2008). A second explanation is that we have only analysed median total soil nitrogen, whereas symbiosis stability in legumes might correlate with plant-available nitrogen and particularly with its lowest values, which fluctuates through the seasons (Cain *et al.*, 1999; Lipson *et al.*, 1999; Scogings *et al.*, 2015). Third, a potential limitation of this analysis is that the scale of resolution of

the soil databases we analysed (1/2 a degree) might not be sufficient to detect variation in soil properties at a fine spatial scale. Although this smaller-scale variation is unlikely to affect global-scale patterns in soil properties, it might obscure relationships between soil variables at a finer scale and symbiotic persistence. A fourth explanation is that the soil environment data is a reflection, rather than a driver, of symbiotic persistence. The lack of correlation with low-quality soil and high symbiotic persistence could be an indication of the historic colonization of the environment by N_2 -fixing legumes, resulting in an influx of nitrogen to the soil (Ledgard & Steele, 1992; Hiltbrunner *et al.*, 2014). This relationship illustrates a limitation of correlating particular habitat variables with evolutionary states: it can be difficult to determine cause and effect. This temporal problem can potentially be addressed by modelling evolution of the nitrogen fixation and soil variables simultaneously in more advanced phylogenetic frameworks (Pagel *et al.*, 2004; Felsenstein, 2012; Kutsukake & Innan, 2014; *e.g.* Revell, 2014), although evolutionary mapping of extrinsic traits like soil properties has attracted criticism (Grandcolas *et al.*, 2011).

Lastly, we found that leaf nitrogen and phosphorus contents were both positively correlated with highly persistent symbioses (Figures 4 and 5). High-N demanding lifestyle and the capacity to engage in N_2 -fixation are thought to be linked (McKey, 1994; Vitousek *et al.*, 2002), our results are consistent with this idea and suggest that it also extends to symbiosis stability. Symbiotic N_2 -fixation allows for a more consistent source of nitrogen and this link potentially drove the evolution of a high nitrogen lifestyle in some legumes, in turn favouring the symbiosis to persist. The positive correlation between leaf phosphorus content and persistent symbiosis is likewise expected because N_2 -fixation is linked with high phosphorus demand as N_2 -fixing efficiency is decreased by low phosphorus availability (Vitousek & Howarth, 1991; Vitousek *et al.*, 2002). Another idea is that symbiotic N_2 -fixation may help alleviate phosphorus limitation because it allows for the secretion of more nitrogen-rich phosphatases, favouring plant phosphorus harvesting (Houlton *et al.*, 2008). Consequentially, high host nutrient demands may increase benefits of being in a symbiotic relationship (Werner *et al.*, 2014c; Wyatt *et al.*, 2014), selecting for persistence of symbiotic capacity. It is still an open question as to whether a high nutrient lifestyle predated symbiosis persistence or the reverse. More detailed research is needed to determine the evolutionary order and causality of the correlation between persistence of nodulation capacity and high nutrient demands. Because a high-nutrient lifestyle is linked with fast growth and more investment in reproduction, that suite of traits is highly beneficial within a particular post-disturbance regeneration niche (*sensu* Grubb, 1977). Legume clades experiencing such a selective regime may have effectively ‘locked’ themselves into a symbiotic lifestyle.

A major question emerging from this research is why facultative symbioses, such as the N_2 -fixing symbiosis, are lost at all? In obligate symbioses, the absence of the symbiotic partner, or a single mutation disrupting the interaction, could immediately lead to extinction of the obligate partner. But why would a facultative species lose the *capacity* for symbiosis? In general, we would expect that three conditions need to be met: (i) an extended period in which the host plant does not need the symbiosis, for instance because the same benefits are provided by the environment, (ii) one or more mutations occur that cause the loss, which may be more or less likely in different clades, and (iii) the species needs to survive potential competition with other legumes that do retain the capacity for symbiosis, and then even perform better if conditions return in which being symbiotic is favourable. Thus, the capacity for symbiosis is lost if there is a long-term selective environment for loss (conditions i and iii) and genetic variation driving symbiosis loss (conditions ii). If there is a cost, small or large, associated with maintaining the underlying genetic mechanism of N_2 -fixation capacity, this can contribute to a selective environment for loss of symbiotic capacity. While there is some theoretical support for this idea (Menge *et al.*, 2008; Lahti *et al.*, 2009), it would be interesting to explicitly test for costs of maintaining facultative symbioses across diverse systems. Similarly, other factors we have considered, like high host nutrient demands, also affect the selective environment, while factors like high molecular rates of evolution (Wright *et al.*, 2006; Gillman *et al.*, 2009) and genome duplication (Li *et al.*, 2013) impact upon the likelihood that mutations disrupt symbiotic capacity. By duplicating genes or whole-genomes, redundant copies of key genes regulating the symbiosis can be created (Young *et al.*, 2011; Li *et al.*, 2013). Because mutations are less likely to disrupt key symbiosis genes, this duplication could result in a lower rate of symbiosis loss and higher retention of symbiotic capacity (Werner *et al.*, 2014a; Cannon *et al.*, 2014). The potential link between symbiosis persistence and gene duplication is an interesting research avenue.

While our aim is to illustrate how quantitative phylogenetic methods can be used to explore patterns of symbiosis evolution, we also recognize limitations to this approach. For example, our dataset of leaf nutrient contents was relatively small compared to our legume occurrence, temperature and soil datasets (Methods). Additionally, these quantitative analyses are only correlative, and thus can only suggest patterns. Experiments are needed to compare the physiology of legumes and other hosts showing different likelihoods of symbiotic persistence (Figure 1), allowing us to test for actual mechanisms that influence symbiosis stability (*e.g.* Behm *et al.*, 2014). For instance, while there has been short-term evolutionary studies analysing the mechanisms of symbiosis loss in rhizobia (Sachs *et al.*, 2010, 2011a), similar work studying the molecular mechanisms underlying nodulation loss in legumes is lacking. Lastly, our work is not a comprehensive analysis of all the factors that could be involved in symbiotic persistence in N_2 -fixation. Rather, it illustrates the type of questions and variables that can be addressed. Other

factors, such as historic atmospheric CO₂-concentrations (Rothman, 2002), have been thought to affect the evolution and stability of other plant root mutualisms (Field *et al.*, 2012, 2015) and could likewise be considered in this framework.

The real power of a quantitative comparative approach is the ability to study evolutionary dynamics across a diversity of symbiosis types, from terrestrial (Bennett & Moran, 2013; Koga & Moran, 2014) to marine habitats (Petersen *et al.*, 2012) in various stages of persistence. For example, some squid species have a very intimate association with bioluminescent bacteria (Kremer *et al.*, 2014). Similar to the legume-rhizobial symbiosis, the bacterial symbionts retain a free-living stage and do not become physically integrated with the host. However, unlike legume hosts that are often found without rhizobial partners (Sprent, 2009), adult squid hosts show a more consistent symbiotic engagement strategy with their bioluminescent symbionts (Koch *et al.*, 2014). Why do these patterns differ and how is this trend reflected in the phylogenetic reconstruction of the partnerships? For instance, do we see fewer instances of loss for symbiotic capacity in squid lineages than legumes and why? Similarly, many insects have bacterial endosymbionts, some of which have been estimated to be associated with the same host clades for millions of years, and have become physically, genetically and metabolically integrated with hosts (Husnik *et al.*, 2013; Bennett & Moran, 2013). One idea is that such integration is favored when symbionts access and deliver new forms of energy for their hosts, rather than simply increasing the efficiency of resource acquisition (Kiers & West, 2015), as happens in the rhizobia-legume symbiosis. This idea could be tested in a comparative framework, asking what host or symbiont traits are associated with symbiont integration versus facultative associations? The first step to integration may be a switch from a more labile symbiotic state to a more persistent state, which can be phenotypically indistinguishable, but may be detected using phylogenetic comparative methods. Understanding these signals, and the general features that promote symbiosis evolution, will require quantitative reconstructions across a diversity of partnerships. Such comparative studies will shed light on how symbioses have solved the problem of aligning and controlling the interests of two species throughout the history of life (West *et al.*, 2015).

Methods

We previously reconstructed the evolutionary history of N₂-fixation using an HRM approach and a database of the nodulation status of 3,467 angiosperm species (Werner *et al.*, 2014a). Our subsequent analyses focused on the 1,366 legumes (Fabaceae) from our database to determine their stable fixing likelihoods, and test possible correlations between stable fixing likelihood and environmental and trait factors. We first confirmed our previous conclusion that nodulating legumes could be subdivided into stable fixers, with an increased symbiotic persistence, and in regular fixers, and that this finding also

holds when analyzing only the legumes. We then tested for correlations between stable fixing likelihood in legumes capable of forming nodules (1,195 species: this subset is smaller because not all 1,366 legumes species can nodulate) (Werner *et al.*, 2014a) and (i) geographic, (ii) climate, (iii) environmental variables and (iv) host traits. We obtained (i) location records for these species from the Global Biodiversity Information Facility (GBIF), (ii) temperature records for these locations from the Climate Research Unit (CRU) (New *et al.*, 2002), (iii) soil clay content and pH values for all legume occurrences from the Harmonized World Soil Database (Nachtergaele *et al.*, 2012) and soil nitrogen from the ISRIC-WISE database (Batjes, 1995, 2012) and (iv) plant leaf nitrogen and phosphorus concentrations from the TRY Plant Trait Database (Kattge *et al.*, 2011) and from the Yasuni Forest Dynamics database (Kraft & Ackerly, 2010).

We used the R-package *rgbif* (Chamberlain *et al.*, 2014) to obtain location records for the nodulating legume species recorded in GBIF and our database, resulting in coverage of 1,182 species. We limited the maximum total number of locations per species to 100,000 (Chamberlain *et al.*, 2014). This procedure resulted in an analysis of almost 3.7 million locations. Subsequently, we scrubbed our location data to exclude all data points with spatial error codes (cddiv, cdout, cdrepf, cdreps, cucdmis, gdativ, preneglat, preneglon, preswcd, zeroed) and obtained a dataset of over 3.2 million legume locations across the globe (median 289 locations per species). We characterised species as stable fixers if they had a corrected stable fixing likelihood of more than 50% (Figures 1 and 2), and otherwise as regular fixers for visualization purposes. We plotted these locations to the world map to identify general geographic patterns (Figure 3), and found a strong latitudinal pattern. We calculated the median latitude per species across all their locations in our dataset and tested if there was a latitudinal pattern in stable fixing likelihood by generating a phylogenetic linear regression model of stable fixing likelihood on absolute median latitude in the R-package *phylolm*, using Pagel's lambda (Pagel, 1999; Tung Ho & Ané, 2014). In order to correct for phylogenetic non-independence, we pruned the full angiosperm phylogeny (Zanne *et al.*, 2013, 2014) to only include those legumes for which we had both median latitude and stable fixing likelihoods.

We then obtained mean long-term annual temperature (TMP_{Mean}) for all locations in our cleaned location dataset, using a bilinear interpolation (Hijmans, 2014) to sample for the location from the climate grid (New *et al.*, 2002). We calculated the median species-level TMP_{Mean} across all locations for each species, only including species for which we had temperature records that covered at least ten different locations, resulting in a coverage of 1,095 legumes. We then explored the extent to which the observed latitudinal pattern was driven by temperature. As previously, we generated phylogenetic linear regression models of stable fixing likelihood as a response variable and its association with absolute median latitude and median TMP_{Mean} in *phylolm* using Pagel's lambda (Pagel, 1999;

Tung Ho & Ané, 2014). We used AIC criteria to establish that the best model contained only temperature as an explanatory variable.

In order to analyse the correlation between stable fixing likelihoods and soil properties, we used a similar approach and employed a bilinear interpolation to sample for the legume occurrence locations, which were then used to obtain soil clay content, soil pH, and total soil N levels from the Harmonized World Soil Database and the ISRIC-WISE database (both $\frac{1}{2}$ a degree resolution) (Batjes, 1995, 2012; Nachtergaele *et al.*, 2012). These values were used to calculate species-level median values. We then generated phylogenetic linear regression models to explain the response variable stable fixing likelihoods in terms of these three soil variables. Using AIC-criteria, we found that inclusion of none of these variables or their combinations provided a better fit than a model that only included an intercept.

We then studied the relationship between symbiotic persistence and host plant traits in nodulating legumes, we obtained a database of plant leaf nitrogen and phosphorus contents from the TRY Plant Trait Database (Fitter & Peat, 1994; Niinemets, 2001; Cornelissen *et al.*, 2003, 2004; Wright *et al.*, 2004; Han *et al.*, 2005, 2012; Louault *et al.*, 2005; Kerkhoff *et al.*, 2006; Preston *et al.*, 2006; Garnier *et al.*, 2007; Craven *et al.*, 2007; Domingues *et al.*, 2007, 2010; Reich *et al.*, 2008, 2009; Fyllas *et al.*, 2009; Kattge *et al.*, 2009, 2011; Baraloto *et al.*, 2010; Ordoñez *et al.*, 2010; Penuelas *et al.*, 2010; Blonder *et al.*, 2011; Milla & Reich, 2011; Butterfield & Briggs, 2011; Prentice *et al.*, 2011; Campetella *et al.*, 2011; Spasojevic & Suding, 2012; Vergutz *et al.*, 2012; Chen *et al.*, 2013) and from the Yasuni Forest Dynamics database (Kraft & Ackerly, 2010). We calculated the mean nutrient content per species for both nitrogen (N_{mass} %) and phosphorus (P_{mass} %) and compiled a dataset of 131 symbiotic legumes for which we have both nitrogen and phosphorus data. Here, we found using AIC that the best phylogenetic linear model explaining stable fixing likelihood among legumes was one that included plant leaf phosphorus and nitrogen content as explanatory variables, as well as their interaction. We normalised N_{mass} and P_{mass} by scaling them by their respective standard deviations so that regression coefficient can be compared between variables.

A potential limitation to the phylogenetic linear model analyses presented here is that we are taking the output of one phylogenetic analysis (*i.e.* stable fixing likelihoods (Werner *et al.*, 2014a)) and using it as an input for subsequent analysis. As defined here, symbiotic persistence can only be discovered in a phylogenetic reconstruction, not directly measured or observed in an organism, necessitating this approach. However, we can verify our results by testing for the phylogenetic correlation between the variables of interest for our best fit models and the underlying binary data on presence or absence

of nitrogen fixation capacity in legumes (Werner *et al.*, 2014b). We therefore generated phylogenetic generalized linear models in *phylolm* (Ives & Garland, 2010; Tung Ho & Ané, 2014) with host legume nodulation capacity as a dependent binary variable, and either mean annual temperature (TMP_{Mean}) or leaf nitrogen (N_{mass} %) and phosphorus (P_{mass} %) content and their interactions as explanatory variables. As previously, we scaled N_{mass} and P_{mass} by their respective standard deviations so that regression coefficient can be compared between variables. We expect that if there are relationships between these variables and stable fixing, they should also be recovered when analysing the correlation between them and fixation per se. As expected, we recovered the same relationship of leaf nitrogen (coefficient: 1.22, $p < 0.01$) and phosphorus content (coefficient 1.11, $p < 0.01$) and their interaction (coefficient -0.37, $p < 0.01$) with symbiotic N_2 -fixation. Additionally, as expected, we recover that legumes at lower TMP_{Mean} are more likely to be able to form the symbiosis (coefficient -0.10, $p = 0.24$). However, this association is not significant, indicating that this result is less robust than the link between symbiotic nitrogen fixation and host nutrient requirements.

Acknowledgements

We are grateful to W.F. Doolittle, P. Keeling, J. McCutcheon and the Sackler Colloquium, LM Henry, L.A.A. Werner, J.I. Sprent and two anonymous reviewers for providing feedback on our manuscript. This research was supported by the Netherlands Organisation for Scientific Research grants 836.10.001 and 864.10.005 to E.T.K. and the European Research Council ERC Grant agreement number 335542 to E.T.K. The study has been supported by the TRY initiative on plant traits (<http://www.try-db.org>). The TRY initiative and database is hosted, developed and maintained by J. Kattge and G. Bönisch (Max Planck Institute for Biogeochemistry, Jena, Germany). TRY is currently supported by DIVERSITAS/Future Earth and the German Centre for Integrative Biodiversity Research (iDiv) Halle--Jena--Leipzig.